The purpose of this study is to examine the differences in reading achievement across a period of three years, as measured by both formative online assessment and state standardized testing, between students participating in online differentiated reading instruction, Unique Reader, and those who did not. Unique Reader and its accompanying Diagnostic Online Reading Assessment (DORA) were adopted district-wide by Hampton City Schools (VA) as part of an effort to maintain consistent vertical literacy testing and record-keeping, and to supplement instruction for struggling students (i.e. defined as those not achieving proficient levels on state testing). The main research questions addressed included: (1) How does elementary student reading achievement change when students are provided supplemental reading instruction in an online environment in addition to regular classroom instruction, and (2) To what extent is this growth reflected in state standardized reading measures, the Virginia Standards of Learning tests (SOL)? Student reading growth was measured throughout each school year via a formative online assessment, DORA; further student academic reading achievement was measured annually using state-mandated testing, the SOL.

Reading educators have long debated the most effective means for teaching students to read. While this debate continues, the structure of effective reading intervention programs often follows a model of individual differentiated instruction based on formative diagnostic data (Juel, 1996; Shanahan & Barr, 1995; Wasik, 1998; McCallum, Roche-Smith & Martin, 2000). Research has further indicated that, under a constructivist theory of reading, obtaining the diagnostic student data upon which to build instruction from multiple sources creates the most complete profile of a student’s reading ability (Flores, Tefft & Esteban, 1991). One such example is the three-cueing system (Adams, 1998) which suggests that the full spectrum of reading can be broken into three overlapping cueing systems: the syntactic, semantic, and grapho-phonics systems.

Based largely on the above theory, Unique Reader (Let’s Go Learn, Inc.) is an online differentiated reading instruction program, driven by diagnostic assessment data generated by Diagnostic Online Reading Assessment (DORA), a formative, criterion-referenced reading assessment. Providing as many as 1000 lessons across four tracks of instruction, Sight Words, Phonics, Vocabulary, and Comprehension, Unique Reader is built on an adaptive-logic platform, and adapts continuously to student achievement.

For this study, we collected students’ DORA data for two consecutive years from a grade-level cohort from Hampton City Schools. We also collected the students state standards-based assessment test scores from the prior year and the two years of DORA testing. We used these scores to address three key research questions:
1. Was there a relationship between students’ DORA and state test-score growth over the study duration, and did students’ DORA scores increase significantly over time?

2. Did students who participated in Unique Reader or who had more DORA testing have greater SOL growth rates?

3. Did students who participated in Unique Reader have greater DORA growth rates?

Information presented in this study can provide practical implications to district-wide implementation of supplemental reading instruction in an online environment.

Method

Data Source

Data from a mid-sized, Southeastern urban school district were collected (N = 1,034) from 2006-2008, during the third-, fourth-, and fifth-grade school years of the participants. Students’ third-grade test scores from spring 2006 represented the first data point. SOL scores also were obtained for each student from spring 2007 and spring 2008 while students were in Grades 4 and 5, respectively. While students were in fourth grade, DORA testing began in the winter. On an as-needed basis, determined by current-year classroom teachers, students were tested again with DORA in the spring of fourth grade, and during the fall, winter, or spring periods when students were in Grade 5. Thus, students had five opportunities over two years to take the DORA assessments. Furthermore, some students were placed into the Unique Reader supplemental instruction program for treatment (n=120). Determination of student placement into the Unique Reader program was made by individual classroom teachers.

Measures

DORA tests across seven subtests: high frequency words, phonics, phonemic awareness, word recognition, oral vocabulary, spelling, comprehension. With the exception of the spelling subtest which is a generative test, all test items are multiple-choice, with either three or four distracters. DORA results are returned as grade-level equivalencies (i.e. a score of 3.6 is equivalent to third grade, six months).

The high frequency word subtest assesses words from Edward B. Fry’s 300 sight words which have been arranged into three general levels of difficulty (Fry, Kress, & Fountoukidis, 2004). The phonics subtest assesses a child's ability to recognize basic English phonetic principles of high utility (Pressley & Woloshyn, 1995), including: 1) beginning sounds, 2) short vowel sounds, 3) blends, 4) the silent E rule, 5) consonant digraphs, 6) vowel digraphs, 7) r-controlled vowels, 8) diphthongs, and 9) syllabification. The phonemic awareness subtest assesses using oral and picture-only items. Categories tested include: 1) addition, 2) deletion, 3) substitution, 4) identification, 5) categorization, 6) blending, 7) segmenting, 8) isolation, and 9) rhyming. The word recognition subtest assesses the student’s ability to recognize words from lists of increasing difficulty. The vocabulary subtest assesses student’s oral vocabulary, using visual definitions. The spelling subtest asks students to generate correct spellings of words of increasing difficulty, based on 1) number of syllables in a word, 2) regular phonetic patterns within the words, 3) irregular phonetic patterns within the words, 4) vocabulary level, and 5) a
child’s expected familiarity with a word based on his or her grade level. Finally, the
comprehension subtest, in a variation on protocols for some informal reading inventories
(Gillet & Temple, 1994; Leslie & Caldwell, 1994), children silently read expository
passages of increasing difficulty and answer questions about each passage immediately
after they read it. The questions for each include three factual questions, two inferential
question, and one contextual vocabulary question.

Those students who participated in the Unique Reader by teacher selection not
only received classroom instruction, but also received online supplemental reading
instruction. Instruction levels were determined by DORA. In addition to DORA
assessment data, Unique Reader usage reports indicate the amount of time each student
spent annually on the supplemental instruction. The time, however, spent on instruction
was also determined by teacher selection.

The state standards-based reading assessment, the SOL, administered each spring,
yielded a single, scaled reading score for each student each year. The state scores were
based on scale that ranged from 0-600, with score of 400 set as the cut-point for
proficiency each year. At each grade level, students are assessed using 30-50 multiple
choice items, developed to assess student knowledge of grade-level indicators, identified
as the Standards of Learning for that particular grade level. The tests across grades were
to not vertically aligned, so the scales from each year were not connected to indicate
across-grade growth.

**Statistical Analysis**

After reviewing the DORA data, we decided to include Word Recognition (WR),
Vocabulary, Spelling, and Comprehension scores in the analyses. These subtests did not
have ceiling effects between grades 5 and 6. We conducted hierarchical linear modeling
(HLM) to address each research question. HLM was particularly suited for the data
because students had multiple scores over time. In the analyses, each student was
considered a Level-2 cluster, and scores over time were nested within each student
cluster. To answer the first research question pertaining to the time-varying covariation
between DORA and SOL scores, the HLM model for each DORA subtest was:

**Level-1 Model**

\[ Y_{it} = \pi_0 + \pi_1 \text{ (Time)} + \pi_2 \text{ (State Score)} + e_{iti} \]

**Level-2 Model**

\[ \pi_0 = \beta_{00} + r_0 \]

\[ \pi_1 = \beta_{10} + r_1 \]

\[ \pi_2 = \beta_{20} + r_2 \]

where \( Y_{it} \) represents the \( i \)th student’s score at time \( t \), Time (coded 0 for the first DORA
scores, and 2 for the last DORA value) and State Scores are Level 1 predictors, and \( e_{iti} \)
indicates individual student error. The \( \pi_1 \) and \( \pi_2 \) values represent each student’s DORA
growth over time and relation with SOL scores over time, respectively. The \( \pi_0 \) reflected
each student’s DORA score at Time 0, or in the winter of fourth grade. In the Level 2
models, the \( \beta_{10} \) and \( \beta_{20} \) indicate the average student DORA growth and DORA-state test
relationship among students. \( \beta_{00} \) indicated the average student DORA score at Time 0.
To address the second research question, the SOL scores were used as the outcome, and Time served as the sole level-one predictor. At Level Two, we included the number of Unique Reader lessons received by each student (coded 0 for students that never received Unique Reader), the number of DORA scores present for each student over two years, and whether or not the student received Unique Reader (coded 1 for Unique Reader students) to predict students’ initial SOL status ($\pi_0$) and growth over time ($\pi_1$). In these models, Time was coded 0 for spring 2006, 1 for spring 2007, and 2 for spring 2008.

**Level-1 Model**

$$Y_{it} = \pi_0 + \pi_1 \text{TIME} + e_{it}$$

**Level-2 Model**

$$\pi_0 = \beta_{00} + \beta_{01}(\text{Unique Reader Lessons}) + \beta_{02}(\text{DORA Test Points}) + \beta_{03}(\text{Unique Reader}) + r_0$$

$$\pi_1 = \beta_{10} + \beta_{11}(\text{Unique Reader Lessons}) + \beta_{12}(\text{DORA Test Points}) + \beta_{13}(\text{Unique Reader}) + r_1$$

For the third question, each DORA subtest served as the outcome in each HLM model. The level-one model included Time as a predictor, and the level-two model included number of Unique Reader lessons and Unique Reader status as predictors of the first DORA score and growth rate over time. The HLM model for each DORA subtest was:

**Level-1 Model**

$$Y_{it} = \pi_0 + \pi_1 \text{TIME} + e_{it}$$

**Level-2 Model**

$$\pi_0 = \beta_{00} + \beta_{01}(\text{Unique Reader Lessons}) + \beta_{02}(\text{Unique Reader}) + r_0$$

$$\pi_1 = \beta_{10} + \beta_{11}(\text{Unique Reader Lessons}) + \beta_{12}(\text{Unique Reader}) + r_1$$

We did not include the number of DORA test scores present per student because many students were missing values at various time points. Thus, scores for these missing time points were imputed using a regression method that involved predicting a missing score at a time point with all other available scores for a given subtest. For the third analysis, we included only those students with at least three of five DORA scores present.

**Results**

HLM models were generated to address the three research questions. The results are presented by question.

1. *Was there a relationship between students’ DORA and state test-score growth over the study duration, and did students’ DORA scores increase significantly over time?*

The findings pertaining to the covariation between DORA and the SOL scores and DORA growth rates over time are presented in Table 1. Each column provides the results from one HLM analysis (one per DORA subtest). The first row in the table provides the estimated average student DORA scores at Time 0, the standard error for these estimates
(in parentheses), and whether the estimates were significantly greater than zero. As can be seen, all initial DORA average scores across the four subtests were significantly greater than zero. The second row of Table 1 includes the estimated average student growth for each subtest, along with the standard errors and the statistical test indicating if each growth rate was significant. As can be seen, students, on average, grew 0.13 score points each study year on the DORA Word Recognition test, which did not indicate a significant gain.

Table 1. *HLM Models for SOL and Time Predicting DORA subtests*

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Word Recognition</th>
<th>Vocabulary</th>
<th>Spelling</th>
<th>Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Initial Status, $\pi_0$</td>
<td>Intercept, $\beta_{00}$</td>
<td>7.69(0.98) **</td>
<td>4.25(0.12) **</td>
<td>2.40(0.10) **</td>
</tr>
<tr>
<td>Model for Time Slope, $\pi_1$</td>
<td>Intercept, $\beta_{10}$</td>
<td>0.13(0.28)</td>
<td>0.28(0.06) **</td>
<td>0.35(0.05) **</td>
</tr>
<tr>
<td>Model for State Test Slope, $\pi_2$</td>
<td>Intercept, $\beta_{20}$</td>
<td>0.003(.002)</td>
<td>0.003(.001) **</td>
<td>0.002(.001)</td>
</tr>
</tbody>
</table>

*Note.* **$p$**<.01, *$p$*<.05.

The growth rates for Vocabulary, Spelling, and Comprehension, however, were significantly greater than zero. The third row in the table contains the DORA and SOL covariation results. As can be seen, students gain in DORA over time covaried positively and significantly with their SOL gain on all subtest except Word Recognition.

2. *Did students who participated in Unique Reader or who had more DORA testing have greater SOL growth rates?*

Table 2 provides the results for the effectiveness of Unique Reader and DORA testing on students’ SOL change over the study duration. The model for the

Table 2. *HLM Model for DORA Test Points & Unique Reader Predicting SOL*

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>SOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Intercept (Initial Status), $\pi_0$</td>
<td>Intercept, $\beta_{00}$</td>
</tr>
</tbody>
</table>
UR Lessons, $\beta_{01}$  
-0.07(0.08)

DORA Test Points, $\beta_{02}$  
-27.17(2.70)**

Unique Reader, $\beta_{03}$  
-14.96(7.21)*

Model for Test Time Slope, $\pi_1$  

Intercept, $\beta_{10}$  
3.99(1.07)**

UR Lessons, $\beta_{11}$  
0.02(0.04)

DORA Test Points, $\beta_{12}$  
2.21(0.97)*

Unique Reader, $\beta_{13}$  
-3.23(3.44)

Note. **$p<.01$, *$p<.05$.

The model for the time slope indicates the degree to which the predictors accounted for variability across students in their rate of SOL change. As can be seen from the table, the Unique Reader variables did not significantly predict change over time, but the number of DORA tests taken by students was a significant and positive predictor. That is, the more students took DORA, the greater their gain in SOL scores over two years. For every additional time students took DORA, test scores, on average, rose 2.21 points on the state test.

3. Did students who participated in Unique Reader have greater DORA growth rates?

For the final research question, four separate HLM models were created (one per DORA subtest) with the two Unique Reader variables serving as potential predictors. Only students with at least three of the possible five DORA scores were included in these
analyses. The two missing time points were imputed for each student based on the three present values. Table 3 provides the results from the analyses. As can be seen in the table,

Table 3. HLM Models for Unique Reader Variables Predicting DORA Subtests

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Word Recognition</th>
<th>Vocabulary</th>
<th>Spelling</th>
<th>Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for Initial Status $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\beta_{00}$</td>
<td>9.61(0.60)**</td>
<td>4.39(0.14)**</td>
<td>2.38(0.12)**</td>
<td>2.63(0.38)**</td>
</tr>
<tr>
<td>UR Lessons, $\beta_{01}$</td>
<td>0.00(0.01)</td>
<td>0.00(0.02)</td>
<td>0.00(0.00)</td>
<td>0.00(0.00)</td>
</tr>
<tr>
<td>Unique Reader, $\beta_{02}$</td>
<td>-1.51(0.89)</td>
<td>-0.34(0.28)</td>
<td>-0.06(0.17)</td>
<td>0.04(0.49)</td>
</tr>
<tr>
<td>Model for Time Slope, $\pi_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\beta_{10}$</td>
<td>-0.68(0.50)**</td>
<td>0.32(0.10)**</td>
<td>0.48(0.10)**</td>
<td>1.13(0.16)**</td>
</tr>
<tr>
<td>UR Lessons, $\beta_{11}$</td>
<td>0.00(0.00)</td>
<td>0.00(0.00)</td>
<td>0.00(0.00)</td>
<td>0.00(0.00)</td>
</tr>
<tr>
<td>Unique Reader, $\beta_{12}$</td>
<td>1.40(0.63)*</td>
<td>0.06(0.13)</td>
<td>-0.07(0.12)</td>
<td>-0.60(0.35)</td>
</tr>
</tbody>
</table>

Note. **p<.01, *p<.05.

students who participated in Unique Reader, or who had more lessons, did not have significantly different initial test scores (winter of fourth grade). Students’ DORA subtest growth rates did not differ based on the number of Unique Reader lessons received, and for Vocabulary, Spelling, and Comprehension, participation in Unique Reader had no effect on change over time. Unique Reader students, however, did have larger Word Recognition growth rates, on average, than students who did not participate in the tutorial modules.

Discussion

The DORA scores for Comprehension, Spelling, and Vocabulary were positively and significantly correlated to state testing scores, indicating that these subtests are demonstrating a correlated growth in students reading to the state testing. It is not surprising to see that the Word Recognition subtest scores were not correlated to SOL scores, as the testing of word identification skills out of context is ordinarily not a skill
that is addressed by most standardized reading assessments, especially those above the
third grade level, at which time, the reading of words in context has taken a much greater
instructional emphasis.

It was also found that students with more DORA assessments had greater gains on
state testing over the two years of the study. This indicates a positive correlation between
the use of formative online assessment and reading achievement. It is possible that this
could be attributed to the amount of diagnostic data provided by DORA results for both
individual and classroom instructional use. It is also possible that students with more
frequent DORA assessments were receiving additional instruction based on the DORA
testing results, and being monitored, using DORA, more closely than their peers. The
implication here is that the students who were being tested more frequently or more
regularly were targeted to receive additional instructional support, either through the use
of Unique Reader, or through other classroom or supplemental programs.

While it was not found that Unique Reader usage was a statistically significant
predictor of change in state testing scores over time, it is possible that this is a function
also of group membership. Since Unique Reader status negatively predicted initial state
test scores, this indicates that the students who were selected for participation in
Unique Reader were already scoring, on average, 15 points behind their peers. As this
was a quasi-experimental study, and students were selected for their participation in
Unique Reader based on teacher selection due to outstanding need, it is possible that the
students in Unique Reader were identified as those students “at risk,” whose normal
academic growth would not have been comparable to their peers to begin with.

The third finding indicated that Unique Reader usage appeared to have a positive
effect on Word Recognition scores, indicating that students’ ability to recognize words
out of context was increased by participating in the Unique Reader program. It is
interesting to note that Unique Reader usage did not have statistically significant effects
on additional subtests, such as comprehension. It is possible that this could be attributed
to the level of Unique Reader usage of the participating students. At lower instructional
levels, Unique Reader focuses on sight words and phonics, skills which would directly
impact students’ word recognition abilities. While ultimately increased comprehension is
the goal of supplemental reading instruction, for students reading below grade level,
instruction often first focuses on decoding skills, which would be seen by increased Word
Recognition scores.

There are limitations in this study to consider. First of all, the quasi-experimental
nature of the study’s design creates comparison groups that are not random in nature and
do not necessarily represent comparable reading abilities in its students. Since students
were selected for participation in Unique Reader based on “need” as assessed by teachers,
it can be assumed that these students were perhaps lower achieving students to begin with
than the students in the control group. Therefore, the growth of the students in the
experimental group would not be expected to be as great at the students who were not
selected to receive additional online instruction, as they were not perceived as being
“needy.”
A further limitation of the study is the unavailability of demographic data on the students. It is unknown, for instance, which students in either the experimental group or the control group were identified special needs or ELL students, characteristics that may very well explain additional variance in both DORA scores and state testing scores. Additionally, demographics such as SES might have also been included to further explain variance in those scores.

Since this study was limited to one grade level in its sample, further studies could examine the same growth over multiple grade levels to determine whether Unique Reader usage serves as a predictor of SOL scores across a larger sample. Additionally, an older sample of students might be used to determine whether subtests like comprehension or vocabulary are correlated to Unique Reader usage, when the program is targeting the needs of more advanced readers. Further, the use of a true experimental study, with random sampling might provide a more widely generalizable prediction of SOL scores from Unique Reader usage as well as DORA testing.

One of the most notable findings of the study, the indication that students who were assessed more frequently using DORA, could very well open the door to further research. Interesting questions include how teachers are using DORA to drive both individual supplemental instruction and to differentiate classroom instruction to meet the needs of those students. Also interesting to investigate are the effects of properly training teachers in the use and interpretation of online diagnostic testing score for use in instructional planning.
References


